Identification of chemical features that influence mycomembrane permeation and antitubercular activity

Irene Lepori[1]*, Zichen Liu[2]*, Nelson Evbarunegbe[3]*, Shasha Feng[4,5]*†, Turner P. Brown[4,5], Kishor Mane[6], Shivangi[6], Mitchell Wong[1], Amir George[6], Taijie Guo[7], Jiajia Dong[7], Joel S. Freundlich[6#], Wonpil Im[4,5#], Anna G. Green[3#], Marcos M. Pires[2#], M. Sloan Siegrist[1,8#]

[1]Department of Microbiology, University of Massachusetts, Amherst; Amherst, MA, USA.
[2]Department of Chemistry, University of Virginia; Charlottesville, VA, USA.
[3]Manning College of Information and Computer Sciences, University of Massachusetts, Amherst; Amherst, MA, USA.
[4]Department of Biological Sciences, Lehigh University; Bethlehem, PA, USA.
[5]Department of Bioengineering, Lehigh University; Bethlehem, PA, USA.
[6]Department of Pharmacology, Physiology and Neuroscience, Rutgers University-New Jersey Medical School; Newark, NJ, USA.
[7]Institute of Translational Medicine, Zhangjiang Institute for Advanced Study, Shanghai Jiao Tong University; Shanghai, China.
[8]Molecular and Cellular Biology Graduate Program, University of Massachusetts, Amherst; Amherst, MA, USA.
†Present address: Merck Research Laboratories, 33 Ave Louis Pasteur, Boston, MA, USA.

*These authors contributed equally to this work
[#]Corresponding authors:
M. Sloan Siegrist[#]
Email: siegrist@umass.edu
Marcos M. Pires[#]
Email: mp7aa@virginia.edu
Anna G. Green[#]
Email: annagreen@umass.edu
Wonpil Im[#]
Email: wonpil@lehigh.edu
Joel S. Freundlich[#]
Email: freundjs@rutgers.edu

Keywords: tuberculosis; machine learning; bioorthogonal click chemistry; permeability; mycomembrane

Abstract

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* (Mtb), is the deadliest single-agent infection worldwide. Current antibiotic treatment for TB takes a minimum of four months, underscoring the need for better therapeutics. The unique mycobacterial cell envelope, particularly the outermost mycomembrane, has long been thought to promote intrinsic antibiotic resistance by limiting compound entry into Mtb. Understanding chemical features that influence permeation across the mycomembrane may enable more accurate predictions of whole cell anti-Mtb activity, leading to development of more effective antibacterials. Here we query the mycomembrane permeation of over 1500 azide-tagged compounds in live Mtb with the bioorthogonal click chemistry-based assay PAC-MAN. We use cheminformatics and machine learning to identify chemical features associated with mycomembrane permeation and show that they have predictive value via systematic modification of two small molecule series. Additionally, we find that chemical features that influence mycomembrane permeation correlate with anti-Mtb activity in large compound libraries. These findings suggest that the mycomembrane is indeed a significant barrier to whole cell activity in Mtb and provide a rational framework for designing or modifying compounds to overcome this barrier.

Introduction

Tuberculosis (TB) caused by *Mycobacterium tuberculosis* (Mtb) consistently claims over one million lives each year[1]. Antibiotic treatment is lengthy, complex, and variably effective. High doses are often required for TB treatment, which in turn increase the risk of side effects and reduce patient compliance. Better drugs and drug regimens to target TB are urgent public health priorities.

An outstanding challenge in developing or improving drugs for TB, as with many other diseases, is ensuring that drug candidates accumulate within the cells implicated in disease pathogenesis[2, 3]. To achieve whole cell activity, a molecule must overcome several barriers, including membranes, efflux, and metabolism, to accumulate with the kinetics and at the concentrations required to effectively engage its target. The physicochemical features or moieties of a molecule that enable it to overcome individual accumulation barriers are divergent and sometimes at odds[4]. Moreover, even within bacteria there are likely inter-species and genera differences in molecule accumulation. For example, the extent and/or pattern of accumulation for certain antibacterials, nutrients, and other molecules differ in mycobacteria relative to Gram-negative bacteria[5-8]. This inference has been generalized with the determination of molecular correlates of accumulation for the Gram-negative species *Escherichia coli* and *Pseudomonas aeruginosa*[9, 10] and the mycobacterial pathogen *Mycobacterium abscessus*[11]. Strikingly, physicochemical properties alone, including those associated with *E. coli* or *P. aeruginosa* accumulation, fail to predict *M. abscessus* accumulation and necessitated a deep learning approach to identify correlates[11]. These findings indicate that molecule accumulation in Mtb is likely to be a complex phenotype.

Identifying chemical features that promote Mtb cell accumulation may enable more rapid development of new tuberculosis treatments. One way to address the complexity of accumulation is to focus on chemical features that enable a molecule to overcome a single barrier. For example, it has long been assumed that Mtb and other mycobacterial pathogens are intrinsically resistant to certain drugs in part because of the impermeability of their outermost mycomembrane[6, 12, 13]. This assumption is predicated on studies showing that mycomembrane disruptions

sensitize mycobacteria to a subset of antibacterials[6, 7, 12-15]. However, activity is an indirect proxy for uptake; collateral metabolic dysfunction from cell envelope perturbation[16-19] may also impact drug sensitivity.

Here we addressed the complexity of Mtb accumulation by first screening a small molecule library for mycomembrane permeation and then deploying machine learning to discover chemical features that influence this phenotype. We used our recently-developed, click chemistry-based assay Peptidoglycan Accessibility Click-Mediated AssessmeNt (PAC-MAN[7, 20]; **Fig. 1a**) to measure the permeation of over 1500 small molecules across the mycomembrane of live Mtb. Cheminformatics analyses suggest that the rules for mycomembrane permeability are not simple, as physicochemical properties have scaffold-dependent effects. To capture structure-function relationships more holistically, we trained a neural network to predict mycomembrane permeability for any compound, then queried for chemical features predictive of permeability. *De novo* chemical synthesis and mycomembrane permeation testing of a two small molecule series confirmed that the chemical features identified by cheminformatics and machine learning are predictive. Finally, we showed that scaffolds and other chemical features that predict mycomembrane permeation also correlate with anti-Mtb activity. Our work supports the long-standing hypothesis that the mycomembrane is a major barrier to whole cell activity in Mtb and suggests that improving the ability of a molecule to permeate the mycomembrane may aid the rational design or redesign of more effective drugs for tuberculosis.

Results

High-throughput screening of mycomembrane permeability

The PAC-MAN assay (**Fig. 1a**) first proceeds via metabolic labeling of peptidoglycan, the cell wall polymer immediately beneath the arabinogalactan-mycomembrane layer of the cell envelope, with a dibenzocyclooctyne (DBCO) probe. A portion of the peptidoglycan-embedded DBCO is captured via bioorthogonal, strain-promoted alkyne-azide cycloaddition (SPAAC[21-23]) with azide-tagged test molecules. The remaining DBCO groups are revealed via SPAAC with a fluorescent azide. The level of mycobacterial cell fluorescence, measured by flow cytometry, inversely correlates with permeation of the azide test molecule across the mycomembrane[7, 20]. We used PAC-MAN to screen Mtb mc^{2}6206 (H37Rv Δ*panCD* Δ*leuCD*)[24] and the model organism *Mycobacterium smegmatis* mc^{2}155 (Msm)[25] with 1572 azide test molecules from three sources: 1152 synthesized via fluorosulfuryl azide chemistry from the corresponding primary amine (Sharpless/Dong[26]), 380 purchased from Enamine, and 40 purchased from various commercial sources (**Fig. 1b**; **Fig. S1**). The latter two sources contain molecules with primary amines. To control for the intrinsic reactivities of azide test molecules toward DBCO[7, 20], we also screened DBCO-functionalized polystyrene beads that have bacteria-like dimensions but are devoid of a permeability barrier (**Fig. 1b**). Previously we controlled for reactivity and calculated mycomembrane permeation by exposing DBCO-labeled beads and mycobacteria to different concentrations of azide test molecules, then calculating the difference in concentration of test azide molecule that competes fluorescence by 50% in the two systems ($\Delta\log_{10}CC_{50}$)[7]. For higher-throughput calculation of

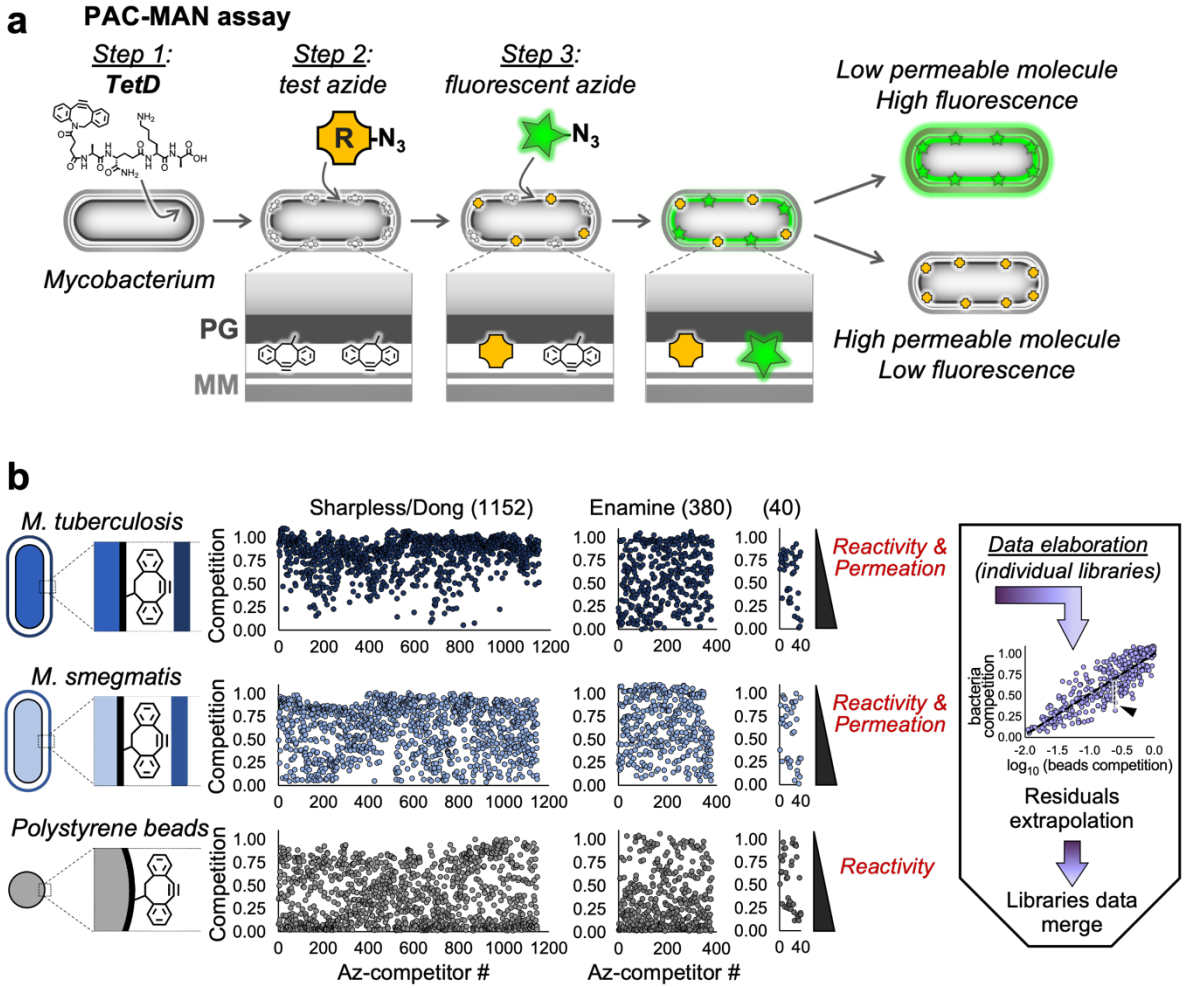

Figure 1. High-throughput screening of mycomembrane permeability. (**a**) Schematic of PAC-MAN assay (adapted from[7]). In *Step 1*, Mtb or Msm is incubated with the DBCO-bearing **TetD** probe to tag cell wall peptidoglycan. Mycobacteria are then exposed to an azide-modified test molecule (*Step 2*) followed by an azide fluorophore (*Step 3*). Test azides that do not permeate the mycomembrane do not access cell wall-embedded DBCO. DBCO that do not react in *Step 2* are free to react with azide fluorophores in *Step 3*, resulting in high fluorescence. Conversely, test azides that permeate the mycomembrane can access and react with DBCO in *Step 2*, preventing DBCO from reacting with azide fluorophores in *Step 3* and resulting in low fluorescence. (**b**) PAC-MAN screening of three test azide libraries: Sharpless/Dong (1152), Enamine (380), or other commercial sources (40). Screening was performed on Mtb, Msm, and DBCO-polystyrene beads to normalize for test azide reactivity. Standardized residuals were extracted from log-linear regression analyses for individual libraries as shown. Values for standardized residuals are inversely proportional to mycomembrane permeability of test azide.

mycomembrane permeation, we instead exposed DBCO-labeled beads and mycobacteria to fixed concentrations of test azide molecules then performed log-linear regression analyses to calculate the differences between observed and

expected fluorescence (standardized residuals; **Fig. 1b**; **Fig. S2**). The standardized residuals for compounds shared between the Sharpless/Dong and Enamine libraries were consistent (**Fig. S3**).

Cheminformatics analyses identify chemical features associated with mycomembrane permeability

We first analyzed the correlation between the presence of small chemical scaffolds and mycomembrane permeability. We used a permissive strategy (**Fig. 2a**, **Fig. S4**), which permits the scaffold of interest to be fused or not with other rings, as well as a "greedy" strategy that excludes the ring-fused versions of the scaffold (**Fig. 2b**). For both Msm and Mtb we found that aromatic nitrogen-containing scaffolds like indole, imidazole, or pyrazole correlate positively with permeation (negative medians for standardized residuals), while scaffolds like cyclopentane or cyclohexane correlate negatively (positive medians).

We next investigated the correlation between mycomembrane permeation and physicochemical properties (**Fig. 2c**). When analyzed as a whole, our Mtb dataset did not reveal obvious correlations. Notably, the physicochemical properties previously shown to correlate with Gram-negative accumulation[9, 10, 27] were not associated with mycomembrane permeation (**Fig. S5**). Reasoning that the contributions of physicochemical properties may depend on the structural context in which they occur, we repeated the cheminformatics analyses after grouping compounds by the presence of specific scaffolds as in **Fig. 2a**. We found that the

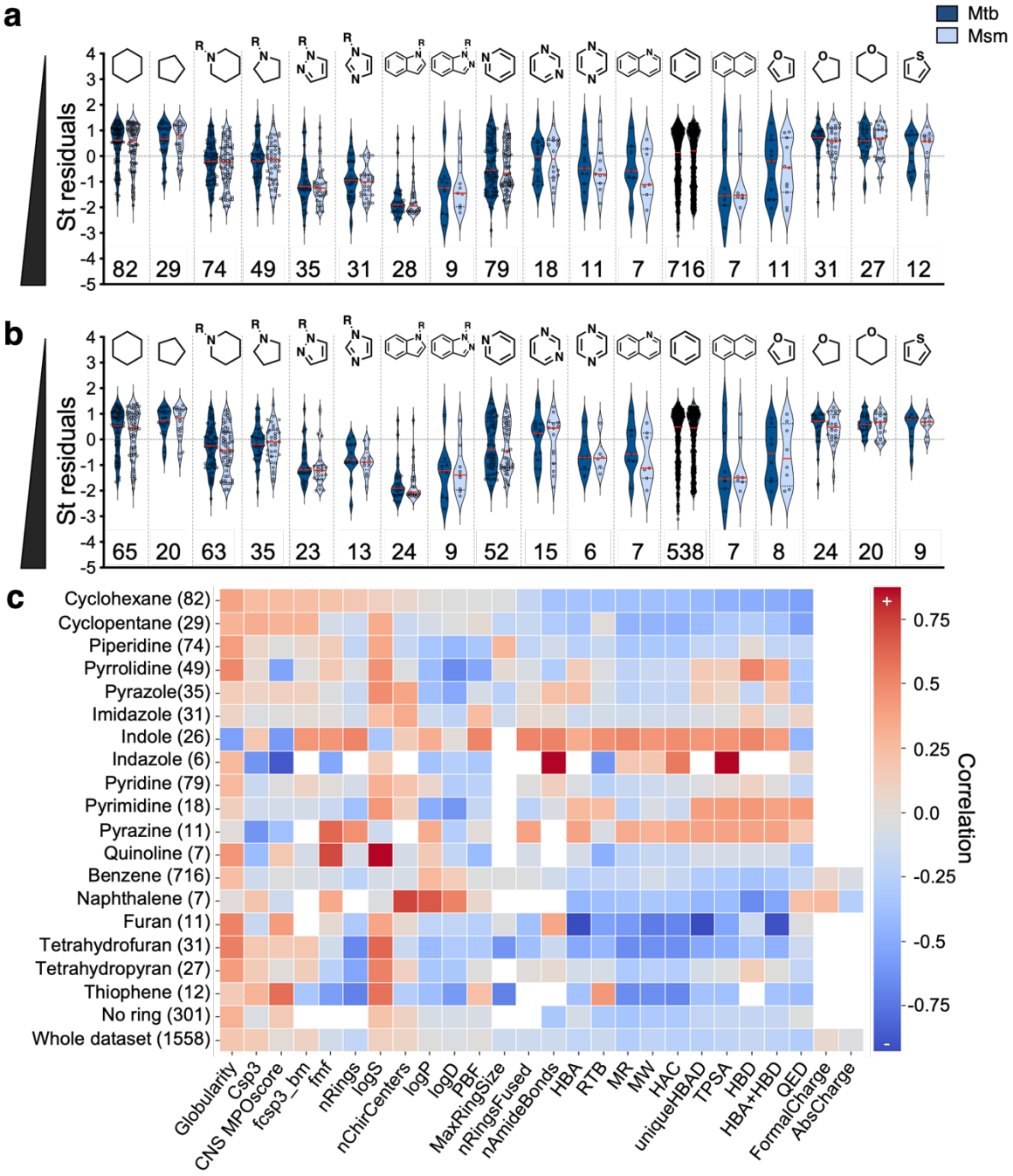

Figure 2. Chemical scaffolds and physicochemical properties that correlate with mycomembrane permeability. (**a**) Mycomembrane permeability (standardized [St] residuals) for test azides grouped by the presence of a specific ring-containing scaffold, indicated on top. (**b**) "Greedy" scaffold analysis excludes the fused rings structures. Numbers below the plots are the number of compounds in each group. (**c**) Correlation between mycomembrane permeability and selected physicochemical properties for all test azides (whole dataset, *bottom*) and for subgroups bearing the scaffolds plotted in (**a-b**). Numbers in brackets are the number of compounds in each group. Red and blue respectively denote positive and negative correlations between the indicated physicochemical properties and mycomembrane permeation. For descriptors explanation see **Figure S1**.

effects of physicochemical properties on mycomembrane permeability varied by scaffold. For example, topological polar surface area (TPSA) and log partition coefficient (logP; Crippen logP calculation from RDKit[28]) have strong, positive correlations with mycomembrane permeation when a molecule contains an indazole or naphthalene, respectively, but have weak and/or negative correlations with mycomembrane permeation in the context of most other scaffolds. These observations are significant as lipophilicity is generally viewed as a positive attribute for antitubercular drugs[29] and underscores the challenge of identifying molecular correlates of Mtb accumulation even for a single barrier.

Machine learning predicts chemical features that are associated with mycomembrane permeation

We built a machine learning (ML) model, called Mycobacterial Permeability neural Network (MycoPermeNet), to capture the complex relationship between chemical structure and mycomembrane permeability (**Fig. 3a**, **Fig. S6**). Inspired by recent successes in deep learning for antibacterial discovery[30, 31], the model takes SMILES strings and Mtb screening data (standardized residuals) as inputs, then uses a two-stage process to predict mycomembrane permeability. First, it learns to generate vector representations of chemical compounds, called embeddings, using a message passing neural network on molecular graphs implemented in Chemprop[32]. It then uses a downstream multilayer perceptron to convert embeddings into a permeability prediction (*i.e.,* predicted standardized residuals). The two stages are trained independently using an 80-10-10 train-validate-test split to optimize hyperparameters and perform model selection. The multilayer perceptron performed

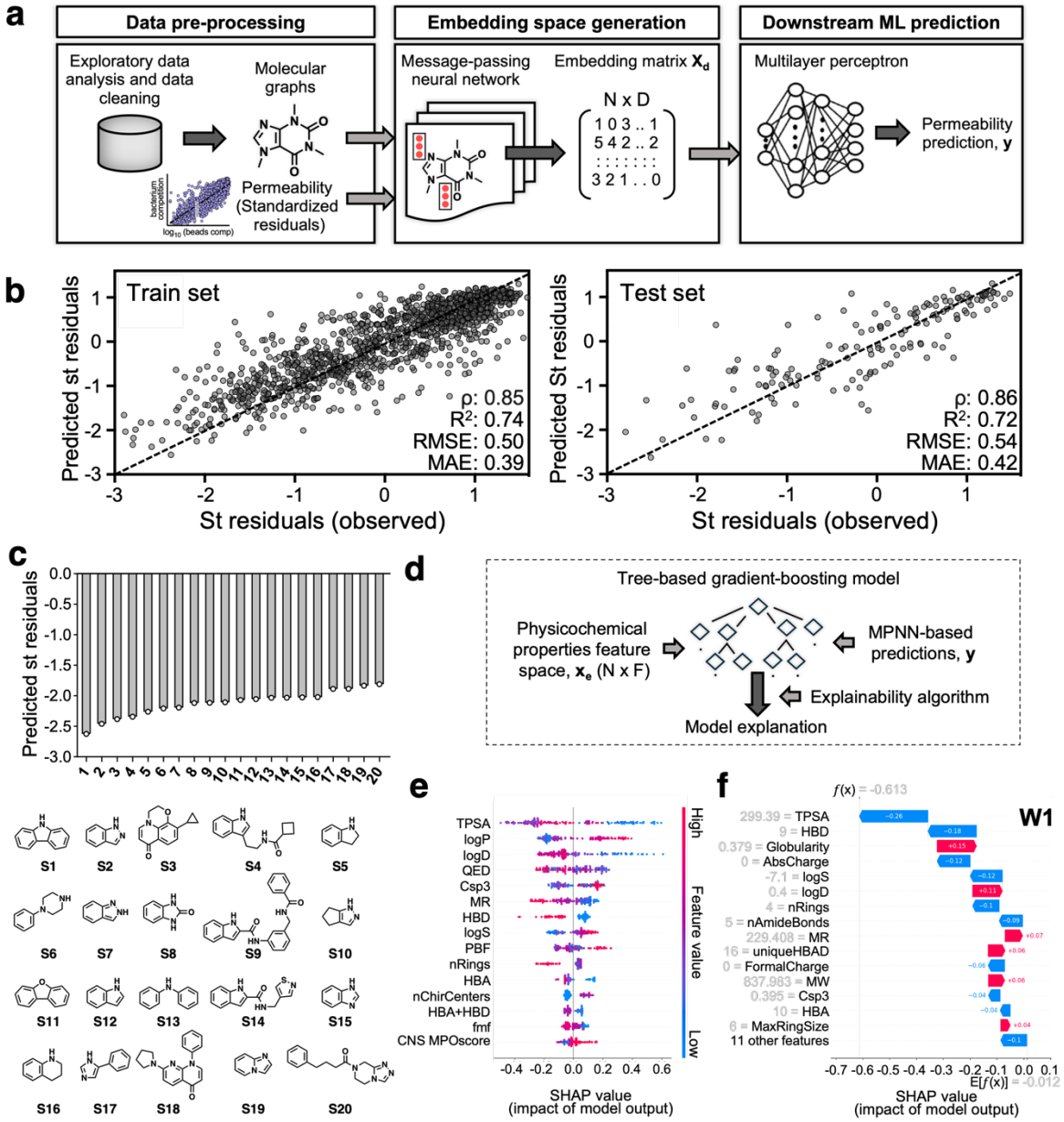

Figure 3. Machine learning model development and interpretation. (**a**) Pipeline for machine learning (ML) model MycoPermeNet development. (**b**) ML-predicted vs. experimentally-observed mycomembrane permeation data (standardized residuals) for both the training (*left*) and validation (*right*) sets. (**c**) The 20 most-permeable chemical scaffolds predicted from the ML model. (**d**) Pipeline for surrogate XGBoost ML model SHAP-based interpretability studies. (**e**) Ranking of top-15 (out of 26) physicochemical properties prioritized by the ML model to drive its prediction across all compounds. Molecular descriptors calculated by RDkit and Collaborative Drug Discovery (CDD) Vault. (**f**) Interpretation of the ML predictions for a specific molecule (**W1** peptide, see **Fig. 4a**) from the perspective of physicochemical properties (descriptors). Rationalization and quantification of the positive or negative effect of the top-15 (out of 26) descriptors on molecule permeability across the mycomembrane. For results relative to the full list of descriptors see **Figure S7**. RMSE: squared root of mean squared error; MAE: mean absolute error. For descriptors explanation see **Fig S1**. N = number of compounds; F = number of descriptors. D = dimension of embeddings.

the best on the validation set out of several model architectures tested (**Fig. S6**). Our final model achieves R^2, RMSE, and MAE respectively of 0.74, 0.50, and 0.37 for the train set and 0.72, 0.37, and 0.41 for the held-out test sets (**Fig. 3b**)**,** indicating a strong relationship between measured and predicted mycomembrane permeability. Moreover, we achieve Spearman rank correlation coefficients of 0.85 on the train set and 0.86 on the held-out test set, demonstrating that our model correctly ranks the relative permeability of compounds even better than it predicts their absolute permeability scores.

To test whether MycoPermeNet is learning reasonable relationships between molecular structure and mycomembrane permeability, we asked the model to predict permeability scores for every Bemis-Murcko scaffold found in the dataset (n=217). Among the 20 scaffolds predicted as most permeable (**Fig. 3c**) we found various indole or indole-like scaffolds, as well as imidazole- and pyrazole-like scaffolds. The concordance between the ML- and cheminformatics-based (**Fig. 2**) analyses serves as a validation of the ML approach and further reinforces that the presence of certain nitrogen aromatic heterocycles correlates with mycomembrane permeability.

To gain additional insights into the physicochemical properties that correlate with molecule permeation across the mycomembrane we performed interpretability studies of MycoPermeNet (**Fig. 3d-f; Fig. S7**). Because our model is based on input chemical structures, which are meaningful individually but difficult to summarize across our large and diverse dataset, we built a surrogate model[33] to determine which human-interpretable chemical descriptors have the highest influence on the permeability predictions made by MycoPermeNet. Specifically, we used a surrogate XGBoost[34] to predict the outputs of MycoPermeNet from 26 hand-selected input

features (23 calculated with RDKit[28] and 3 calculated with Collaborative Drug Discovery [CDD] Vault[35]) that we considered important for mycomembrane permeability. After training our surrogate XGBoost model (R^2, RMSE, and MAE respectively of 0.84, 0.33, and 0.25 for the train set and 0.83, 0.38, and 0.28 for the held-out test set, where the R^2 on the held-out test set explains 83 percent of the variance in relationship between the features and MycoPermeNet), we interpret its features using SHAP[36] (SHapley Additive exPlanations; **Fig. 3d**). In this analysis, the two most influential physicochemical properties are TPSA and logP (**Fig. 3e**; **Fig. S7a**). While SHAP returns properties that drive mycomembrane permeability predictions across the dataset, it can also predict which features are responsible for the permeability prediction for individual compounds (**Fig. 3f**; **Fig. S7b**). This analysis provides both a qualitative and quantitative interpretation of the impact of physicochemical properties on mycomembrane permeability.

Chemical features that are associated with mycomembrane permeation have high predictive value

We hypothesized that molecular features identified by cheminformatics and/or predicted by MycoPermeNet to correlate with mycomembrane permeability are causative. We first tested this hypothesis with a pentapeptide series (Phe-Lys-Phe-Lys-Phe) in which we systematically substituted phenylalanines for tryptophans (**Fig. 4a**). The side chains of phenylalanine and tryptophan respectively bear benzene, which is weakly and negatively associated with mycomembrane permeability, and

232 indole, which is strongly and positively associated with mycomembrane permeability

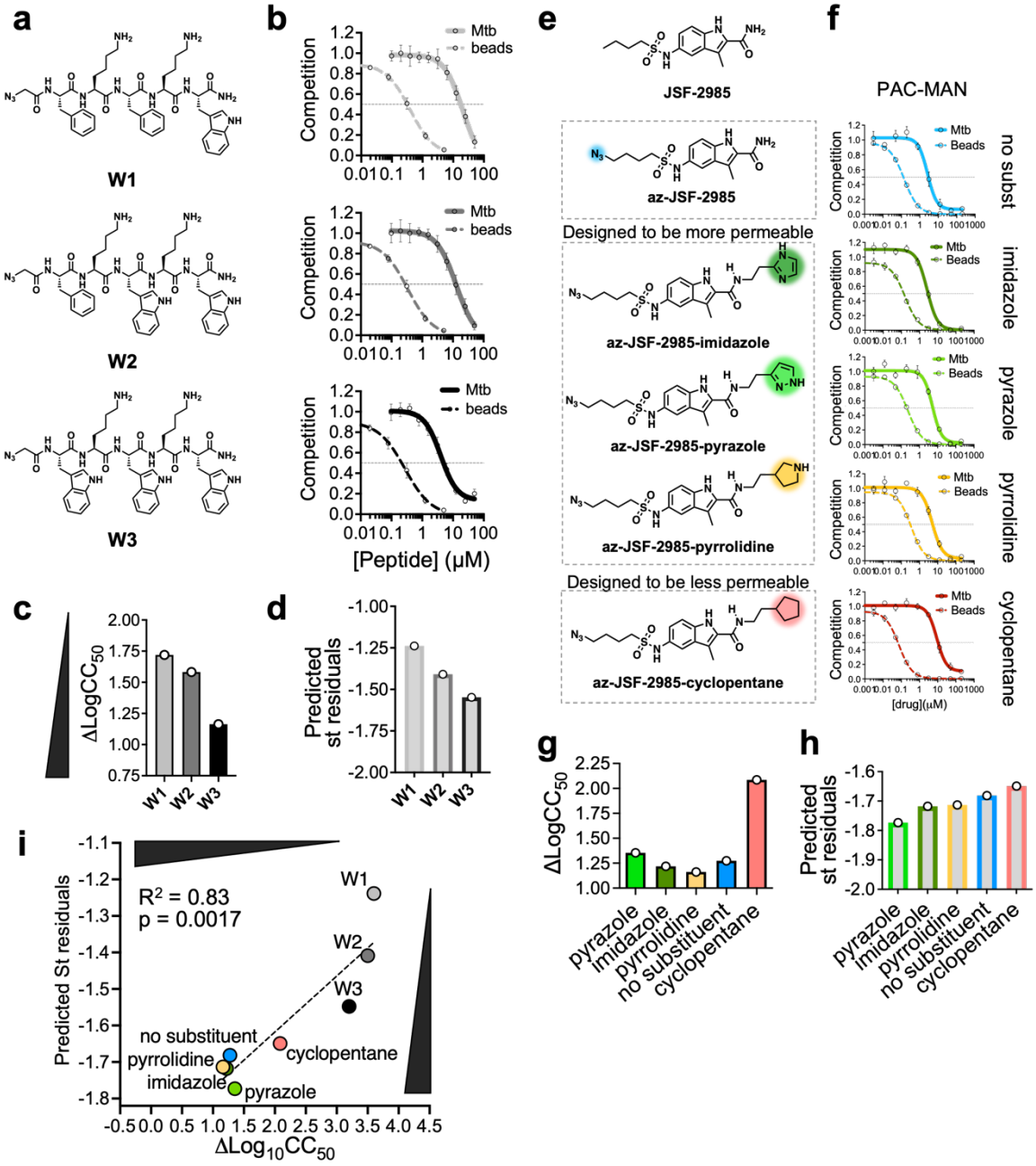

Figure 4. Chemical features identified by cheminformatics and ML influence mycomembrane permeation. (**a-d**) Analysis of a pentapeptide series in which benzene-containing phenylalanines are systematically substituted with indole-containing tryptophans (**W1-3**). The presence of indole or benzene scaffolds are respectively associated with more or less mycomembrane permeability (**Fig. 2a-b**). (**a**) Peptide chemical structures; (**b**) PAC-MAN results for Mtb and beads; (**c**) observed ($\Delta\log_{10}CC_{50}$) and (**d**) ML-predicted mycomembrane permeability. (**e-h**) Analysis of an antitubercular candidate series. **JSF-2985** (**e**, *top*) was derivatized to bear an azide and chemical scaffolds that are associated with more (imidazole, pyrrole, pyrrolidine) or less (cyclopentane) mycomembrane permeability (**Fig. 2a-b**). (**f**) PAC-MAN results for Mtb and beads; (**g**) observed ($\Delta\log_{10}CC_{50}$) and (**h**) ML-predicted mycomembrane permeability. (**i**) Correlation between observed and ML-predicted mycomembrane permeability across the two compound series.

(**Fig. 2a-b**). We found that substitution of phenylalanines for tryptophans enhanced mycomembrane permeation of the peptides in both Mtb and Msm (**Fig. 4b-c**; **Fig. S8**), consistent with both cheminformatics analyses (**Fig. 2a-b**) and MycoPermeNet predictions (**Fig. 4d**).

We next tested our hypothesis with a small molecule series based on **JSF-2985**, an antitubercular small molecule we previously reported[37]. We synthesized **JSF-2985** analogs via replacement of the 2-position primary amide's N-H with moieties that correlate positively (imidazole, pyrazole, and pyrrolidine) or negatively (cyclopentane) with mycomembrane permeation (**Fig. 4e**). We chose the 2-position amide N-H as the location of the new substituents based on docking studies that suggest that the amide group is not directly involved in the engagement of the target but instead lies outside the binding pocket and it is thus exposed to the solvent (**Fig. S9**). We found that **JSF-2985** derivatives with imidazole, pyrazole, and pyrrolidine scaffolds permeate the mycomembrane better than the **JSF-2985** derivative bearing cyclopentane (**Fig. 4f-g**), consistent with both cheminformatics analyses (**Fig. 2a-b**) and MycoPermeNet predictions (**Fig. 4h**). The excellent correlation between ML-predicted and observed permeation of the mycomembrane across the two small molecule series (**Fig. 4i**) suggests the potential to rationally tune the ability of a molecule to traverse this barrier.

Chemical features that predict mycomembrane permeability correlate with whole cell activity

We and many others have shown that mycomembrane disruption sensitizes Mtb to some antibacterials[6, 7, 12-15]. Thus, we hypothesized that mycomembrane permeation

is a key determinant of whole cell anti-Mtb activity. Our docking studies (**Fig. S9**) suggested that the **JSF-2985** derivatives have a high probability of target binding. However, we found that Mtb growth inhibition by these molecules did not reflect their mycomembrane permeation profiles (**Fig. S9b-c**), highlighting the complex relationship between molecule accumulation and activity.

We reasoned that a larger dataset, both in terms of the number and size of molecules, would better test our hypothesized relationship between mycomembrane permeation and activity. To this end, we performed a retrospective analysis on a small molecule library previously screened for anti-Mtb activity[38, 39]. We looked for sets of molecules within the ~200k Molecular Libraries Small Molecule Repository (MLSMR; data obtained from CDD database[35], Burlingame, CA. www.collaborativedrug.com) collection that have similar structures but differ in the presence or absence of mycomembrane permeability-promoting scaffolds (**Fig. 2a-b**) in peripheral positions. We chose this scheme to decrease the likelihood of the scaffold contributing directly to target engagement. In the four distinct molecule sets that met our criteria (**Fig. 5a-b**; **Fig. S10**), compound activity correlated with the presence of these scaffolds and, more generally, with ML-predicted mycomembrane permeation (note that standardized residuals are inverted in **Fig. 5**, **Fig. S10**, and **Fig. S12** for more intuitive comparisons to activity).

276 Because we do not know the targets of these compounds, we next examined the

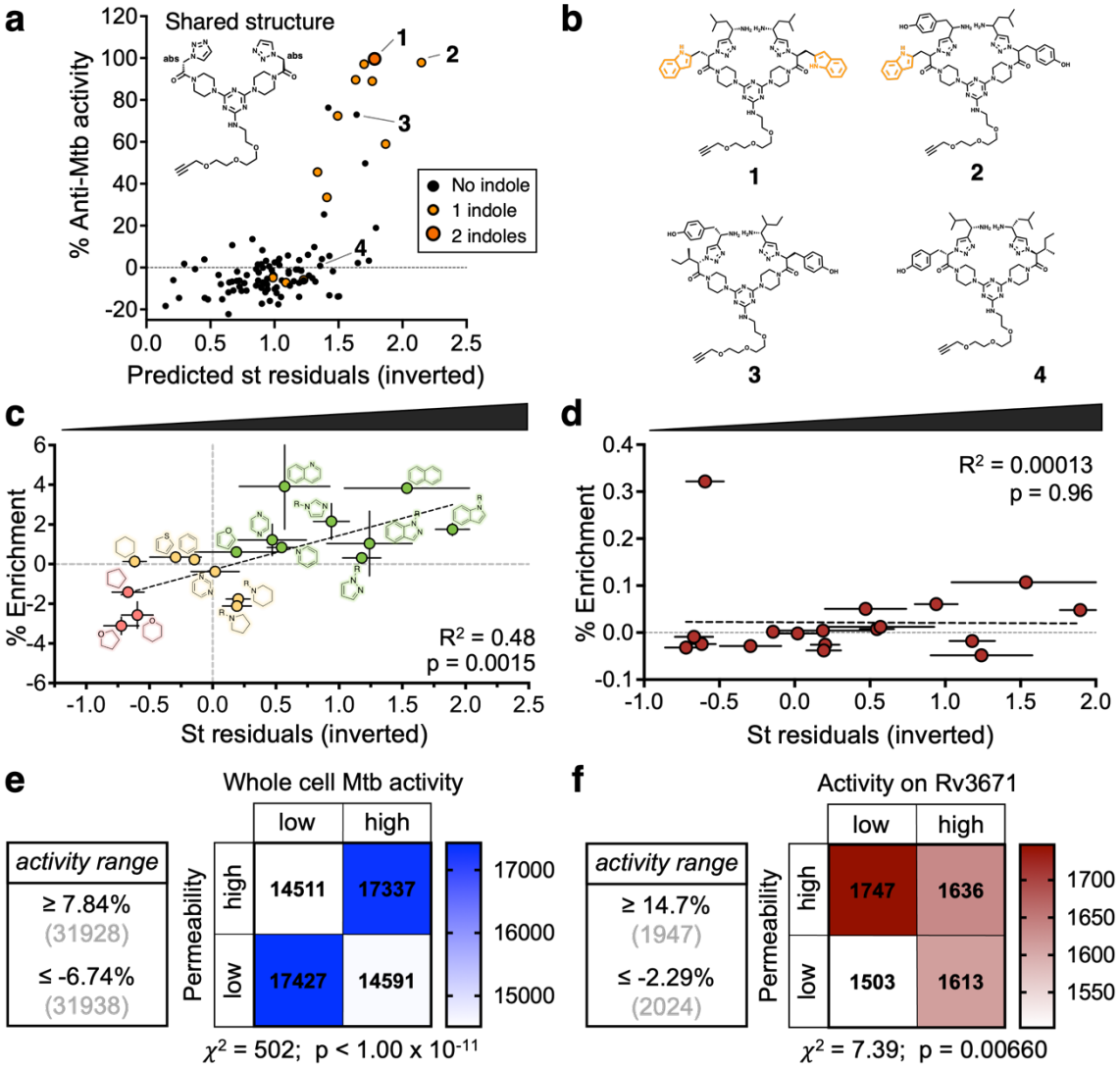

Figure 5. Scaffolds and chemical features that influence mycomembrane permeation correlate with whole cell anti-Mtb activity. (**a**) Relationship between whole cell anti-Mtb activity and ML-predicted mycomembrane permeability for 101 MLSMR compounds that share the indicated structure. The presence of an indole is highlighted in orange. (**b**) Four examples of structures from (**a**), one for each combination +/-indole and +/- activity. (**c-d**) Scaffold-specific relationship between activity against whole Mtb cells (**c**) or against a purified Mtb enzyme (serine protease Rv3671c (**d**)) and observed mycomembrane permeation (inverted standardized residuals). X axis is median mycomembrane permeability for test azides that share indicated scaffolds (from **Fig. 2a**). Y axis is percentage of activity enrichment for MLSMR and TAACF molecules that share the same scaffold relative to all the molecules in the respective datasets (**Figure S12**). Y axis data available from CDD (**c**) or publicly available from PubChem (AID 2606) (**d**). (**e-f**) Relationship between (**e**) whole cell anti-Mtb activity (MLSMR) or (**f**) purified Rv3671c enzyme inhibition and ML-predicted mycomembrane permeation. Each library was divided into quadrants (threshold for permeability: median; threshold for activity: bottom and top 15% of the total values of the libraries (MLSMR) or bottom and top 1% values (Rv3671c), see **Figure S12** for further analyses) and the number of values per quadrant were plotted in (**e**) and (**f**) respectively. The association in (**e**) is significantly stronger than the association in (**f**) by Cramer's V.

277 relationship between chemical structure, mycomembrane permeability, and whole
278 cell activity more broadly. We widened our analysis to three different screens: the
279 MLSMR (above) and Tuberculosis Antimicrobial Acquisition and Coordinating Facility
280 (TAACF; data obtained from CDD[35]) small molecule collections, which were
281 screened against whole Mtb cells, and a third screen against the purified Mtb
282 enzyme Rv3671c (PubChem AID 2606). In addition to activity data, the libraries
283 afforded us the opportunity to examine molecules that do not have azide tags and
284 occupy broader chemical space than that covered by our screening libraries (**Fig.**
285 **S11**). We found that molecule activity in the whole cell screens, but not the enzyme
286 screen, correlates with both observed permeability of different chemical scaffolds
287 (**Fig. 5c-d**; **Fig. S12**) and with ML-predicted mycomembrane permeability (**Fig. 5e-f**;
288 **Fig. S12**). These data suggest that scaffolds and other chemical features that
289 influence mycomembrane permeation are correlates of whole cell anti-Mtb activity.

290

291 **Discussion**

292 Mass spectrometry studies that assess whole cell association have found that
293 molecule accumulation varies across *E. coli, P. aeruginosa*, and *Acinetobacter*
294 *baumannii*, likely reflecting species-specific differences in molecule transit via porins,
295 diffusion across the outer membrane, and transit via efflux pumps[9, 10, 40]. Given that
296 the Mtb mycomembrane has a structure[6, 12, 13] and porin-like proteins[41, 42] that are
297 distinct from those of the Gram-negative outer membrane, it is not surprising that the
298 predictors of mycomembrane permeation that we identify here are different from
299 those of Gram-negative bacteria accumulation[9, 10, 27]. Future use of Mtb strains with

targeted disruptions to mycomembrane transporters or structural integrity may enable the identification of pathway-specific predictors of permeation.

One limitation of PAC-MAN is that it requires the presence of an azide. Azides are widely recognized as ideal bioorthogonal tags because they are small and have minimal known impacts on the physicochemical properties of the parent compound[21, 43]. While we cannot rule out potential effects of the azide on mycomembrane permeation, we note that the chemical features that we identified as predictors of this phenotype from azide-tagged libraries are also correlates of whole cell activity for untagged compounds (**Fig. 5**; **Fig. S10**; **Fig. S12**). We speculate that the constancy of the azide across all the test molecules, the large number of test molecules, and the redundancy of some test molecules (*i.e.*, compounds that differ only in the position of the azide) contribute to our ability to predict mycomembrane permeation.

Intracellular accumulation in Mtb depends on the ability of a molecule to overcome membrane, efflux, and metabolism barriers[2, 3]. Because it covalently traps test molecules in the mycobacterial cell wall, PAC-MAN primarily measures mycomembrane permeation, *e.g.*, passive diffusion and facilitated transport[7, 20]. As well, our test molecules are relatively simple in structure as they have a median size of 180 Da and generally abide by Lipinski's rule of five[44]. The restrictions on cell compartment and chemical space queried are both a strength and limitation of our work. On the one hand, we were able to identify clear structure-function relationships for an aspect of intracellular accumulation. Moreover, these relationships, where tested, accurately predicted mycomembrane permeation. Chemical predictors of mycomembrane permeation have obvious potential in informing the (re)design of anti-Mtb therapeutics against periplasmic targets. Direct and indirect data from this

work and the literature, respectively, suggest these predictors may have even broader utility, *i.e*., evaluation of compounds with cytoplasmic targets. On the other hand, the structure-function relationships that we identify here may not suffice as predictors for larger, more complex molecules or for whole cell accumulation. Expansion of the chemical space covered by our azide libraries beyond Lipinski's rule of five[45], along with whole cell mass spectrometry[4, 5, 9-11, 40, 46, 47], adaptation of PAC-MAN for the cytoplasm[48], and the use of strains with defined membrane, efflux, or metabolism defects[4, 48, 49] will collectively enable the generation of more comprehensive accumulation models.

Acknowledgments

We thank the director of the University of Massachusetts Amherst Flow Cytometry facilities, Dr. Amy Burnside, for her help and advice. These studies were supported by the National Institutes of Health (R01 AI179080 to W.I., M.M.P., and M.S.S. as well as U19 AI142731 and R01 AI153145 to JSF), by the University of Massachusetts Amherst Institute for Applied Life Sciences Midigrant and Core Facilities Incentive Funds to I.L., and by the Gates Foundation INV-080847 to M.M.P. and M.S.S.. This work utilized resources from Unity, a collaborative, multi-institutional high-performance computing cluster managed by UMass Amherst Research Computing and Data.

Author contributions

I.L., Z.L., N.E., S.F. equally contributed to this work; I.L., Z.L., N.E., S.F., J.S.F., W.I., A.G.G., M.M.P., and M.S.S. designed research; I.L., Z.L., N.E., S.F., T.P.B., K.M., S., M.W., A.G., and T.G. performed research; I.L., Z.L., N.E., S.F., T.P.B., K.M., S., M.W., A.G., T.G., J.D., J.S.F., W.I., A.G.G., M.M.P., and M.S.S. analyzed data; J.D., J.S.F., W.I., A.G.G., M.M.P., and M.S.S. supervised the research; I.L. and M.S.S. wrote the manuscript with the help of all authors. The manuscript was discussed and approved by all authors.

Competing Interest Statement

M.S.S. is a co-founder and acting CSO of Latde Diagnostics.

Supplementary Information is available for this paper

Materials Availability

All reagents generated in this study are available upon request from the corresponding authors.

Code and Data Availability

All code, processed input data, and saved model weights for the cheminformatics analyses and MycoPermeNet model are available upon request and will be made publicly available prior to final publication on github: https://github.com/Nevbarunegbe/Mycomembrane-permeability-project.

Material and methods

For a full description of the experimental procedures see **Supplemental Methods** file.

References

1. Organization, W. H. *Global tuberculosis report 2024*; 2024.
2. Tanner, L.; Denti, P.; Wiesner, L.; Warner, D. F., Drug permeation and metabolism in Mycobacterium tuberculosis: Prioritising local exposure as essential criterion in new TB drug development. *IUBMB Life* **2018,** *70* (9), 926-937.
3. Awasthi, D.; Freundlich, J. S., Antimycobacterial Metabolism: Illuminating Mycobacterium tuberculosis Biology and Drug Discovery. *Trends Microbiol* **2017,** *25* (9), 756-767.
4. Zhao, S.; Adamiak, J. W.; Bonifay, V.; Mehla, J.; Zgurskaya, H. I.; Tan, D. S., Defining new chemical space for drug penetration into Gram-negative bacteria. *Nat Chem Biol* **2020,** *16* (12), 1293-1302.
5. Davis, T. D.; Gerry, C. J.; Tan, D. S., General platform for systematic quantitative evaluation of small-molecule permeability in bacteria. *ACS Chem Biol* **2014,** *9* (11), 2535-44.
6. Jarlier, V.; Nikaido, H., Mycobacterial cell wall: structure and role in natural resistance to antibiotics. *FEMS Microbiol Lett* **1994,** *123* (1-2), 11-8.
7. Lepori I., J. K., Liu Z., Chordia M.D., Wong M., Rivera S.L., Roncetti M., Poliseno L., Pires M.M., Siegrist M.S., The mycomembrane differentially and heterogeneously restricts antibiotic permeation. *bioRxiv* **2025**.
8. McGowen K., F. T., Wang X., Zinga S., Wolf I.D., Akusobi C.C., Denkinger C.M., Rubin E.J., Sullivan M.R., Efflux pumps and membrane permeability contribute to intrinsic antibiotic resistance in *Mycobacterium abscessus*. *bioRxiv* **2024**.
9. Geddes, E. J.; Gugger, M. K.; Garcia, A.; Chavez, M. G.; Lee, M. R.; Perlmutter, S. J.; Bieniossek, C.; Guasch, L.; Hergenrother, P. J., Porin-independent accumulation in Pseudomonas enables antibiotic discovery. *Nature* **2023,** *624* (7990), 145-153.
10. Richter, M. F.; Drown, B. S.; Riley, A. P.; Garcia, A.; Shirai, T.; Svec, R. L.; Hergenrother, P. J., Predictive compound accumulation rules yield a broad-spectrum antibiotic. *Nature* **2017,** *545* (7654), 299-304.
11. Sullivan M., R. E. J., Deep learning-based prediction of chemical accumulation in a pathogenic mycobacterium. *bioRxiv* **2024**.
12. Brennan, P. J.; Nikaido, H., The envelope of mycobacteria. *Annu Rev Biochem* **1995,** *64*, 29-63.
13. Batt, S. M.; Minnikin, D. E.; Besra, G. S., The thick waxy coat of mycobacteria, a protective layer against antibiotics and the host's immune system. *Biochem J* **2020,** *477* (10), 1983-2006.
14. Li, S.; Poulton, N. C.; Chang, J. S.; Azadian, Z. A.; DeJesus, M. A.; Ruecker, N.; Zimmerman, M. D.; Eckartt, K. A.; Bosch, B.; Engelhart, C. A.; Sullivan, D. F.; Gengenbacher, M.; Dartois, V. A.; Schnappinger, D.; Rock, J. M., CRISPRi chemical genetics and comparative genomics identify genes mediating drug potency in Mycobacterium tuberculosis. *Nat Microbiol* **2022,** *7* (6), 766-779.
15. Xu, W.; DeJesus, M. A.; Rucker, N.; Engelhart, C. A.; Wright, M. G.; Healy, C.; Lin, K.; Wang, R.; Park, S. W.; Ioerger, T. R.; Schnappinger, D.; Ehrt, S., Chemical Genetic Interaction Profiling Reveals Determinants of Intrinsic Antibiotic Resistance in Mycobacterium tuberculosis. *Antimicrob Agents Chemother* **2017,** *61* (12).
16. Koh, E. I.; Oluoch, P. O.; Ruecker, N.; Proulx, M. K.; Soni, V.; Murphy, K. C.; Papavinasasundaram, K.; Reames, C. J.; Trujillo, C.; Zaveri, A.; Zimmerman, M. D.; Aslebagh, R.; Baker, R. E.; Shaffer, S. A.; Guinn, K. M.; Fitzgerald, M.;

Dartois, V.; Ehrt, S.; Hung, D. T.; Ioerger, T. R.; Rubin, E. J.; Rhee, K. Y.; Schnappinger, D.; Sassetti, C. M., Chemical-genetic interaction mapping links carbon metabolism and cell wall structure to tuberculosis drug efficacy. *Proc Natl Acad Sci U S A* **2022,** *119* (15), e2201632119.
17. Cho, H.; Uehara, T.; Bernhardt, T. G., Beta-lactam antibiotics induce a lethal malfunctioning of the bacterial cell wall synthesis machinery. *Cell* **2014,** *159* (6), 1300-11.
18. Pohane, A. A.; Carr, C. R.; Garhyan, J.; Swarts, B. M.; Siegrist, M. S., Trehalose Recycling Promotes Energy-Efficient Biosynthesis of the Mycobacterial Cell Envelope. *mBio* **2021,** *12* (1).
19. Kohanski, M. A.; Dwyer, D. J.; Hayete, B.; Lawrence, C. A.; Collins, J. J., A common mechanism of cellular death induced by bactericidal antibiotics. *Cell* **2007,** *130* (5), 797-810.
20. Liu, Z.; Lepori, I.; Chordia, M. D.; Dalesandro, B. E.; Guo, T.; Dong, J.; Siegrist, M. S.; Pires, M. M., A Metabolic-Tag-Based Method for Assessing the Permeation of Small Molecules Across the Mycomembrane in Live Mycobacteria. *Angew Chem Int Ed Engl* **2023,** *62* (20), e202217777.
21. Agard, N. J.; Prescher, J. A.; Bertozzi, C. R., A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J Am Chem Soc* **2004,** *126* (46), 15046-7.
22. Sletten, E. M.; Bertozzi, C. R., Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew Chem Int Ed Engl* **2009,** *48* (38), 6974-98.
23. Lepori I., O. Y., Im J., Ghosh N., Paul M., Schubert U.S., Fedeli S., Bioorthogonal “Click” Cycloadditions: A Toolkit for Modulating Polymers and Nanostructures in Living Systems. *reactions* **2024,** *5* (1), 231-245.
24. Sampson, S. L.; Dascher, C. C.; Sambandamurthy, V. K.; Russell, R. G.; Jacobs, W. R., Jr.; Bloom, B. R.; Hondalus, M. K., Protection elicited by a double leucine and pantothenate auxotroph of Mycobacterium tuberculosis in guinea pigs. *Infect Immun* **2004,** *72* (5), 3031-7.
25. Snapper, S. B.; Melton, R. E.; Mustafa, S.; Kieser, T.; Jacobs, W. R., Jr., Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis. *Mol Microbiol* **1990,** *4* (11), 1911-9.
26. Meng, G.; Guo, T.; Ma, T.; Zhang, J.; Shen, Y.; Sharpless, K. B.; Dong, J., Modular click chemistry libraries for functional screens using a diazotizing reagent. *Nature* **2019,** *574* (7776), 86-89.
27. Cain, B. N.; Hergenrother, P. J., Using permeation guidelines to design new antibiotics-A PASsagE into Pseudomonas aeruginosa. *Clin Transl Med* **2024,** *14* (3), e1600.
28. G., L. *RDKit: Open-source cheminformatics*, https://www.rdkit.org, 2013.
29. Makarov, V.; Salina, E.; Reynolds, R. C.; Kyaw Zin, P. P.; Ekins, S., Molecule Property Analyses of Active Compounds for Mycobacterium tuberculosis. *J Med Chem* **2020,** *63* (17), 8917-8955.
30. Stokes, J. M.; Yang, K.; Swanson, K.; Jin, W.; Cubillos-Ruiz, A.; Donghia, N. M.; MacNair, C. R.; French, S.; Carfrae, L. A.; Bloom-Ackermann, Z.; Tran, V. M.; Chiappino-Pepe, A.; Badran, A. H.; Andrews, I. W.; Chory, E. J.; Church, G. M.; Brown, E. D.; Jaakkola, T. S.; Barzilay, R.; Collins, J. J., A Deep Learning Approach to Antibiotic Discovery. *Cell* **2020,** *181* (2), 475-483.
31. Wong, F.; Zheng, E. J.; Valeri, J. A.; Donghia, N. M.; Anahtar, M. N.; Omori, S.; Li, A.; Cubillos-Ruiz, A.; Krishnan, A.; Jin, W.; Manson, A. L.; Friedrichs, J.; Helbig, R.; Hajian, B.; Fiejtek, D. K.; Wagner, F. F.; Soutter, H. H.;

Earl, A. M.; Stokes, J. M.; Renner, L. D.; Collins, J. J., Discovery of a structural class of antibiotics with explainable deep learning. *Nature* **2024,** *626* (7997), 177-185.

32. Heid, E.; Greenman, K. P.; Chung, Y.; Li, S. C.; Graff, D. E.; Vermeire, F. H.; Wu, H.; Green, W. H.; McGill, C. J., Chemprop: A Machine Learning Package for Chemical Property Prediction. *J Chem Inf Model* **2024,** *64* (1), 9-17.

33. C., M., *Interpretable Machine Learning: A Guide For Making Black Box Models Explainable*. https://christophm.github.io/interpretable-ml-book/, 2022.

34. Chen T., G. C. In *XGBoost: A Scalable Tree Boosting System*, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA USA, 13 August 2016; Association for Computing Machinery: San Francisco, CA USA, 2016.

35. Hohman, M.; Gregory, K.; Chibale, K.; Smith, P. J.; Ekins, S.; Bunin, B., Novel web-based tools combining chemistry informatics, biology and social networks for drug discovery. *Drug Discov Today* **2009,** *14* (5-6), 261-70.

36. Lundberg S.M., L. S.-I. In *A Unified Approach to Interpreting Model Predictions*, Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA USA, Long Beach, CA USA, 2017.

37. Inoyama, D.; Awasthi, D.; Capodagli, G. C.; Tsotetsi, K.; Sukheja, P.; Zimmerman, M.; Li, S. G.; Jadhav, R.; Russo, R.; Wang, X.; Grady, C.; Richmann, T.; Shrestha, R.; Li, L.; Ahn, Y. M.; Ho Liang, H. P.; Mina, M.; Park, S.; Perlin, D. S.; Connell, N.; Dartois, V.; Alland, D.; Neiditch, M. B.; Kumar, P.; Freundlich, J. S., A Preclinical Candidate Targeting Mycobacterium tuberculosis KasA. *Cell Chem Biol* **2020,** *27* (5), 560-570 e10.

38. Maddry, J. A.; Ananthan, S.; Goldman, R. C.; Hobrath, J. V.; Kwong, C. D.; Maddox, C.; Rasmussen, L.; Reynolds, R. C.; Secrist, J. A., 3rd; Sosa, M. I.; White, E. L.; Zhang, W., Antituberculosis activity of the molecular libraries screening center network library. *Tuberculosis (Edinb)* **2009,** *89* (5), 354-63.

39. Ananthan, S.; Faaleolea, E. R.; Goldman, R. C.; Hobrath, J. V.; Kwong, C. D.; Laughon, B. E.; Maddry, J. A.; Mehta, A.; Rasmussen, L.; Reynolds, R. C.; Secrist, J. A., 3rd; Shindo, N.; Showe, D. N.; Sosa, M. I.; Suling, W. J.; White, E. L., High-throughput screening for inhibitors of Mycobacterium tuberculosis H37Rv. *Tuberculosis (Edinb)* **2009,** *89* (5), 334-53.

40. Leus, I. V.; Adamiak, J.; Chandar, B.; Bonifay, V.; Zhao, S.; Walker, S. S.; Squadroni, B.; Balibar, C. J.; Kinarivala, N.; Standke, L. C.; Voss, H. U.; Tan, D. S.; Rybenkov, V. V.; Zgurskaya, H. I., Functional Diversity of Gram-Negative Permeability Barriers Reflected in Antibacterial Activities and Intracellular Accumulation of Antibiotics. *Antimicrob Agents Chemother* **2023,** *67* (2), e0137722.

41. Wang, Q.; Boshoff, H. I. M.; Harrison, J. R.; Ray, P. C.; Green, S. R.; Wyatt, P. G.; Barry, C. E., 3rd, PE/PPE proteins mediate nutrient transport across the outer membrane of Mycobacterium tuberculosis. *Science* **2020,** *367* (6482), 1147-1151.

42. Korycka-Machala, M.; Pawelczyk, J.; Borowka, P.; Dziadek, B.; Brzostek, A.; Kawka, M.; Bekier, A.; Rykowski, S.; Olejniczak, A. B.; Strapagiel, D.; Witczak, Z.; Dziadek, J., PPE51 Is Involved in the Uptake of Disaccharides by Mycobacterium tuberculosis. *Cells* **2020,** *9* (3).

43. Jewett, J. C.; Bertozzi, C. R., Cu-free click cycloaddition reactions in chemical biology. *Chem Soc Rev* **2010,** *39* (4), 1272-9.

517 44. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., Experimental
518 and computational approaches to estimate solubility and permeability in drug
519 discovery and development settings. *Adv Drug Deliv Rev* **2001,** *46* (1-3), 3-26.
520 45. Dash, R.; Liu, Z.; Lepori, I.; Chordia, M. D.; Ocius, K.; Holsinger, K.;
521 Zhang, H.; Kenyon, R.; Im, W.; Siegrist, M. S.; Pires, M. M., Systematic
522 Determination of the Impact of Structural Edits on Accumulation into Mycobacteria.
523 *bioRxiv* **2025**.
524 46. Wang, X.; Perryman, A. L.; Li, S. G.; Paget, S. D.; Stratton, T. P.;
525 Lemenze, A.; Olson, A. J.; Ekins, S.; Kumar, P.; Freundlich, J. S., Intrabacterial
526 Metabolism Obscures the Successful Prediction of an InhA Inhibitor of
527 Mycobacterium tuberculosis. *ACS Infect Dis* **2019,** *5* (12), 2148-2163.
528 47. Chakraborty, S.; Gruber, T.; Barry, C. E., 3rd; Boshoff, H. I.; Rhee, K. Y.,
529 Para-aminosalicylic acid acts as an alternative substrate of folate metabolism in
530 Mycobacterium tuberculosis. *Science* **2013,** *339* (6115), 88-91.
531 48. Ongwae G.M., L. Z., Chordia M.D., Dalesandro B.E., Guo T., Dong J., Pires
532 M.M., Determination of Accumulation of Molecules in Escherichia coli. *bioRxiv* **2023**.
533 49. Krishnamoorthy, G.; Leus, I. V.; Weeks, J. W.; Wolloscheck, D.; Rybenkov,
534 V. V.; Zgurskaya, H. I., Synergy between Active Efflux and Outer Membrane
535 Diffusion Defines Rules of Antibiotic Permeation into Gram-Negative Bacteria. *mBio*
536 **2017,** *8* (5).

537